

Abstract

A technique for incorporating binary formatting into a tag-based description language, such as XML, is provided. The binary formatting is achieved by tokenizing the tag and attribute names into variable sized numeric tokens, thereby obviating the need for repetitive or redundant storage of lengthy unicode words, etc. The binary formatting minimizes parsing time and the generation of overhead incident to the formatting and parsing of data. Parsing time is thereby substantially decreased and generally, the size of the resulting file decreases too.